Exploratory Team Cognition and Resilience

in Human Agent Teaming

by

Glenn Lematta

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2019 by the
Graduate Supervisory Committee:

Rod Roscoe, Co-Chair
Nancy Cooke, Co-Chair
Erin Chiou

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Human-agent teams (HATs) are expected to play a larger role in future command and control systems where resilience is critical for team effectiveness. The question of how HATs interact to be effective in both normal and unexpected situations is worthy of further examination. Exploratory behaviors are one that way adaptive systems discover opportunities to expand and refine their performance. In this study, team interaction exploration is examined in a HAT composed of a human navigator, human photographer, and a synthetic pilot while they perform a remotely-piloted aerial reconnaissance task. Failures in automation and the synthetic pilot's autonomy were injected throughout ten missions as roadblocks. Teams were clustered by performance into high-, middle-, and low-performing groups. It was hypothesized that high-performing teams would exchange more text-messages containing unique content or sender-recipient combinations than middle- and low-performing teams, and that teams would exchange less unique messages over time. The results indicate that high-performing teams had more unique team interactions than middle-performing teams. Additionally, teams generally had more unique team interactions in missions with novel degraded conditions than in missions without novel degraded conditions. Implications and suggestions for future work are discussed.

ACKNOWLEDGMENTS

This work was partially supported by ONR Award N000141110844 (Program Managers: Marc Steinberg, Paul Bello). I acknowledge the large effort from the team of undergraduate and graduate researchers who aided in data collection and recruitment for this experiment. Finally, the assistance of Steven M. Shope and Paul Jorgenson, Sandia Research Corporation was vital in developing and integrating the testbed.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Autonomous technology is anticipated to span a wide range of application in future command and control systems (Endsley, 2015). However, as autonomous agents advance to fulfill more complex roles, human-machine interdependence shifts to resemble team-like relationships. Research has identified several requirements for autonomous machines to team with people (Groom & Nass, 2007; Klein et al., 2004; Johnson et al., 2014). For instance, an autonomous search and rescue robot may have the capacity to adjust its route planning to the constraints of a collapsed building, but the resulting behavior may be unexpected to human teammates. Coordinating an explanation for this unexpected behavior may require understanding a teammate's expectations (Chakraborti et al., 2019) including what they would find appropriate for an explanation (Miller, 2018), and strategies for providing explanations. Alternatively, an autonomous combat vehicle may maneuver on a battlefield but be less capable of interpreting broad command intent (Woods et al., 2004). The autonomous combat vehicle's rigidity may impose high levels of workload vehicle crews during unexpected events and may also result in counterintuitive vehicle actions. These examples highlight more fundamental challenges in human-agent teams which are different from all-human teams.

A discriminating factor in assessing a team's effectiveness is the ability for the team to coordinate in new ways when perturbed, and return to previous levels of performance (Gorman et al., 2010). Team resilience is demonstrated in response to unexpected perturbations through coordinating a shared understanding of the event, re-planning, and re-establishing common ground (Hoffman & Hancock, 2017). Insight into the behaviors of high-reliability organizations has also led to an appreciation of how teams-level states of resilience may be cultivated proactively (Bowers et al., 2017). Team-level activities such as planning, leadership, and adaptation are all active processes that emerge from the dynamic interaction of teammates in their environments (Cooke et al., 2013). So although synthetic teammates may not be very resilient themselves, human-agent teams may overcome challenges, and perhaps in qualitatively different ways than all-human teams.

Taking a systems perspective suggests that principles of adaptation in human-agent teams may be similar to adaptation in other complex adaptive systems (Demir et al., 2018). Of particular interest in this research is how adaptive systems generate variation and utilize feedback. For instance, genetic variation in combination with environmental feedback explains adaptation in biological systems over time (Caporale & Doyle, 2013, p. 21). In teams, parallels may be identified in the ways that humans vary performance to meet the demands of everyday work (Hollnagel, 2013). Relying on an autonomous agent or human teammate during an unexpected event may be difficult when team interactions. Team interaction exploration may lead to the discovery and refinement of team coordination process (Gorman et al., 2010), and may be required in unexpected situations where process may not be specified in advance (Sarter et al., 1997). Team interaction exploration may also provide insight into how teams actively cultivate states of resilience as well as how teams develop solutions to novel challenges. The purpose of this study is to examine the relationship between team interaction exploration and performance in normal and unexpected situations in an aerial reconnaissance HAT.

CHAPTER 2

BACKGROUND

Human-Agent Teaming

Since the advent of autonomous capabilities, systems designers have struggled with redefining automation’s role in sociotechnical systems. Questions regarding the role of automation have historically been addressed in models of function allocation, such as Fitts’s list (Fitts 1951, p. 10) and levels of automation (Parasuraman et al., 2000). Previous models have framed function allocation problems as understanding an agent’s capacities as functionally independent, while ignoring how agents provide capacity to each other through coordination. Interdependence between agents describes how collectives manage dependencies (Malone & Crowston 1994) which, may be essential for completing a team task or situationally opportunistic (Johnson et al., 2014). Trending approaches to function allocation reject notions that automation can ‘substitute’ work seamlessly and that humans should be taken out of the control loop (Bradshaw et al., 2013). Some notable function allocation frameworks include adaptive automation (Parasuraman et al., 1992), mixed-initiative interaction (Allen et al., 1999), collaborative control (Fong et al., 2001) supervisory control (Sheridan & Parasuraman 2004), adjustable autonomy (Bradshaw et al., 2003), and coactive design (Johnson et al., 2014). Each framework addresses challenges related to introducing automation with complex cognitive capability and interdependence in automation.

Artificial agents with degrees of decision-making capability or *autonomy* have recently become more prevalent in command and control systems (Cox, 2013). Some autonomous agents may also resemble a teammate more than tools or components of a larger automated system. For instance, an autonomous agent could operate one or more remotely-piloted aerial vehicles while coordinating with humans in the loop (Ball et al., 2009; Mercado et al., 2016). The agent's activities constitute a collection of interdependent activities typically ascribed to a team-member role. In these cases, and particularly when autonomous agents possess decision-making authority in several essential tasks, the relationship between human and non-human agents is more appropriately classified as a *human-agent team* (HAT). HAT is actively advancing in command-and-control domains such as human-robot teams in disaster response (Chakraborti et al., 2019; Demir et al., 2018), autonomous personal assistants space exploration teams (Bradshaw et al., 2003), and various military applications (Brewer et al., 2018; McNeese et al., 2018; Endsley, 2015; Mercado et al., 2015).

Demir and Cooke (2018) examined HAT by informing participant dyads that a third co-located teammate was a synthetic agent, when it was in fact a human. This study took place in a remotely-piloted aerial system synthetic task environment (RPAS-STE; see Cooke & Shope 2005), where participants formed teams to complete aerial reconnaissance missions in varied environmental conditions. They found that teams who were informed their third teammate was a synthetic pilot perceived less workload, gave more suggestions, and reported liking the pilot more than when participants were told their teammate was human. Other studies have explored coordination dynamics (Demir et al., 2019), team synchrony (Demir et al., 2017), and team situation awareness (Demir et al., 2017). In a recent study, inexperienced HATs with a synthetic pilot were compared to inexperienced all-human teams (i.e. control) as well as all-human teams lead by an experienced pilot (McNeese et al., 2017). This study analyzed team performance, target processing efficiency, team situation awareness, and team verbal behaviors (e.g. status updates, inquiries) across five missions, and found that teams lead by an experienced pilot performed best, followed by all-human teams and HATs, which were not significantly different from each other. Additionally, synthetic teams requested more information, rather than providing information proactively.

McNeese et al. (2017) noted that the synthetic teammate's lack of anticipation of the informational needs of other teammates seemed to be reciprocated by human teammates. This is similar to Chiou and Lee's (2016) findings in studying human-agent cooperation that participants were less cooperative with uncooperative agents in a resource exchange task. This study used a microworld task environment resembling a hospital scheduling scenario to test resource exchange strategies in human-agent interactions with varying levels of agent cooperativeness and temporal task demands. They found that humans tended to reciprocate the degree of cooperation exhibited by the agent, engaging in less resource sharing when interacting with the less cooperative agent. While coordinating via natural language texting may have more complex interaction dimensions than discrete resource exchanges, these findings taken together suggest that there may be a contagious effect of interaction style between computer agents and other teammates. Overall, HATs demonstrate potential but have shortfalls related to coordination flexibility and anticipation of teammate needs and cooperation.

Team Cognition and Resilience

Teams are preferred in organizations largely because they can adaptively solve problems (Cooke et al., 2013). However, not all problems are created equal. Many problems may be well-adapted to, with pre-defined responses, such as contingency plans and standard operating procedures. By contrast, system complexity leads to a seemingly limitless supply of possible events that may occur unexpectedly. Woods (2018) proposed a theory of graceful extensibility to explain the tradeoff between refinement within a scoped boundary of performance or *optimization* and investing resources into extending capabilities near or at performance boundaries or *resilience*. According to the theory, an adaptive system models its own range of performance within the environment. It uses this model to allocate resources to fulfill specific capacities that increase fitness to the environment. In doing so, the system trades-off between allocating resources to increase fitness within the modeled boundary of performance and allocating resources to define and expand this boundary. Resilient systems extend their capabilities to manage surprising situations.

Several team-level macro-cognitive activities have been associated with team resilience. These include anticipation, monitoring, planning, and learning (Bowers et al., 2017; Hoffman and Hancock, 2017; Hollnagel, 2016). In a model of team resilience posited by Bowers et al., (2017) team resilience is described as a second-order emergent state, which is emergent from the combination of team states including *cohesion*, *collective efficacy*, *culture*, *shared mental models*, *familiarity*, and *adaptability*. Their model provides several processes that may enhance team resilience such as compensatory behavior, adaptability, and performance monitoring. Additionally, extensibility theory suggests that some mechanisms that aid in team resilience may also support the refinement of capabilities, whereas the allocation of resources through mechanisms may fall on an optimality-resilience spectrum (Woods, 2018). In teams, team interactions may lead to the discovery of such opportunities through exploring new ways of interacting.

Team Interaction Exploration

Exploration is a phenomenon observed in search behavior of complex adaptive systems. In a review of exploration, Hills et al., (2015) identified patterns of exploration in human cognition, human behavior, and social problem solving. Specifically, they elaborate the tradeoff between exploration of potential resources and exploitation of resources that have been identified. Exploratory learning has been examined previously in teams (Kostopoulos, & Bozionelos, 2011, Kostopoulos et al., 2013). Kostopoulous & Bozionelos (2011) found that exploratory team learning may lead to the discovery of new capabilities, whereas exploitative learning refines the capabilities that are established within a team. This study developed a multiple-item survey to examine exploratory and exploitative learning in innovation teams. Additionally, they posited that psychological safety or the "shared belief that the team is safe for interpersonal risk taking" (Edmondson, 1999, p. 354) has been associated with exploratory learning in teams (Kostopoulos & Bozionelos, 2011). Psychological safety as a team state changes as a consequence of work system dynamics, as with other team states such collective efficacy (Bowers et al., 2017). As team interactions involving risk taking facilitate teamwork, team psychological safety may increase. Conversely, team interactions may lead to negative outcomes such as increased workload, miscalibrated trust, or reduced performance. Therefore, feedback on the consequences of team interactions plays a critical role in team interaction exploration, allowing for a deeper understanding of teamwork and the work system.

Promoting proactive interaction exploration with advanced autopilots has been suggested as a solution to automation surprise by improving mental models and developing a repertoire of interactions (Sarter et al., 1997). Considerable effort has also been put forth to develop explainable-artificial intelligence systems (Miller, 2018), which effectively allow users to explore system boundaries and calibrate trust in these system (Hoffman et al., 2018). Other instances of exploration include animal foraging (Hills et al., 2012), honeybee scouting (Beekman et al., 2007), and infant behavior (Gibson, 1988, p. 12-19). While research has considered team exploration in teams as a high-level learning strategy, there is a gap in team-level exploratory activities, such as team cognition. The exploration of team cognition may be assessed directly through team interactions (Cooke et al., 2013). Thus, this study defines exploratory team interactions as *any team interaction unique to a team's collective interaction history*. Dimensions of team interactions refer to qualities such as number of recipients, content, pattern, and situation. Note that exploratory team interactions are distinct from team interactions facilitating an exploratory behavior strategy. Rather, exploratory team interactions are characterized by how the system of signals and feedback between teammates is explored in different situations.

Exploring new ways of interacting as a team may have an underlying intention, such as increasing coordination flexibility or identifying a solution to a novel problem. Intentionality may suggest one or more agents are aware of constraints that affect the value of a particular team interaction. However, constraints in the environment may lead to unintentional novel variations in team interactions. For instance, under high-temporal demand, an operator may communicate more hastily and omit some information.

Importantly, team interaction exploration may have immediate effects, such as communicating the same information more efficiently or obtaining status information, and long-term effects (e.g. team-process learning) for both intentional and unintentional exploration.

Team interaction exploration may support team resilience in a number of ways. For example, a team could use low-workload periods to develop and rehearse new contingency plans, elaborate team-level grounding, and participate in collective activities that build trust. There is evidence that teams who were able to coordinate in new ways were more effective in a command and control task with unexpected roadblocks (Gorman et al., 2010). Novel interactions may contribute to team situation awareness, particularly when coordinating the perception and response to off-nominal situations (Gorman et al., 2006). Furthermore, Cooke et al., (2007) noted that "There appears to be a trade-off between training teams for repeated precision in an unchanging environment and training adaptive teams." (p. 152). Dynamical analysis of teamwork has also identified that highly rigid and highly flexible teams were associated with worse team performance and that optimal performance was associated with metastable team coordination (Demir et al., 2019). In dynamical systems, metastable states demonstrate a sustained equilibrium between stable and unstable dynamics rather than drifting toward either highly rigid or highly unstable equilibrium states (Kelso, 2002). For team interactions, more repeated team interaction patterns correspond with more stable dynamics, while more exploratory interaction patterns correspond with less stable dynamics. Team interaction, therefore, is a mechanism for teams to calibrate stability and maintain agility in addition to its primarily role as an active team-level cognitive process.

When considering the role of artificial agents in team interaction exploration, it is noteworthy that current agents have a much more limited capacity to explore new team interactions than humans. As autonomous systems interact with the environment, they may generate novel behaviors through responses to variations in situational constraints. Yet, current artificial agents have a limited capacity to understand novel situations and adapt their coordination over time. This is concerning for HAT co-adaptation as findings in the RPAS-STE showed that HATs developed less flexible interactions than human teams (Demir et al., 2019). Therefore, human adaptation to autonomous agents dynamically influences the agent's interactions, and in turn the agent's responses adjust to changing inputs. Team interaction exploration may be particularly relevant for HATs in mitigating these team dynamics and compensating for synthetic teammates' lack of exploration.

Team interaction exploration may be related to other relevant constructs in HAT, such as team trust, team situation awareness, and workload. Team trust was associated with high-performing HATs, whereas middle- and low-performing teams were less trusting (McNeese et al., 2019). Exploring interactions with a synthetic teammate may improve one's understanding of teamwork with that agent, which may lead to more appropriate trust and reliance. For instance, exploring a request for information from the synthetic agent may indicate that the agent is unable to provide that information. This means that the requestor can find another way to access this information for when it is needed. Agent transparency of an intelligent decision-support agent was also associated with greater performance and trust in a multiple remotely-piloted aerial systems operation task (Mercado et al., 2016).

For each team member to have the appropriate situation awareness to perform in their role and in off-nominal situations, each role should obtain the information they need in a timely manner (Gorman et al., 2006). Measures that focus on team interactions may capture the processes that lead to emergent states of team situation awareness. In the RPAS-STE, HATs that engaged in more 'pushing' than 'pulling' interactions had higher team situation awareness during roadblocks in a command-and-control task (Demir et al., 2017). *Pushing* refers to the sender providing information to a recipient (e.g. status update, suggestion), while a *pulling* refers to the sender requesting information from the recipient. Team interaction exploration facilitates the discovery of ways to push and pull information and may be required for coordinating the perception and response to an unexpected event. Finally, team interaction exploration has relevant workload considerations. On the one hand, that all coordination has associated costs (Hoffman & Woods, 2011) means that exploring a team interaction with a low payoff may impose excessive workload or be distracting. Yet, gaining familiarity through team interaction exploration may reduce excessive workload proactively. Critical in managing coordination costs seems to be identifying what interactions are worth exploring in teams and when to explore them.

CHAPTER 3

PROJECT OVERVIEW

The aim of the current study is to examine team interaction exploration in HATs while working together in a dynamic and complex task environment. To examine the phenomena in a command and control team task, this study asked how exploratory team interactions affect team performance in an aerial reconnaissance task under normal and degraded conditions. Degraded conditions should require novel coordination, while normal conditions benefit from refined routine coordination. Additionally, exploration should become less common over time as relevant interactions for normal situations are discovered and repeated. However, given exploratory team interactions may correspond with novel situations, this effect might not hold for in missions with degraded conditions. The following was hypothesized:

H1: High-performing teams will explore team interactions more than medium- and low-performing teams.

H2: Teams will explore more team interactions in missions with novel degraded conditions than in missions without novel degraded conditions.

This study is part of a larger effort to develop a cognitively plausible synthetic teammate for a three-agent remotely-piloted aerial system (RPAS) ground crew (Ball et al., 2010). The synthetic teammate is fully capable of performance as the pilot role in normal conditions but considers only routine coordination needs. Several experiments were previously conducted to test and validate the synthetic teammate under normal conditions. Because degraded conditions are common in dynamic command and control tasks, the current study is interested in how this human-agent team performs in normal and degraded conditions. The goal of the task was to capture good photos of target waypoints. To accomplish this, teams needed to navigate the RPAS through restricted zones while avoiding hazards (e.g. mountainous terrain or enemy waypoints). In a single mission, there are 11-12 target waypoints in total. Each agent was assigned to a specific role. The *navigator* plans a route and shares waypoint information with the *pilot*, who uses information received from teammates to fly the RPAS. The *photographer* negotiates RPAS settings with the pilot to take a photo and share feedback on progress with the team. The synthetic teammate assigned to the pilot role is capable of predetermined responsibilities, including requesting information from the navigator, maintaining the RPAS settings, and negotiating with the photographer. Teams interacted with each other via text-chat interface.

Seven degraded conditions were injected throughout ten missions to examine teamwork over time and in multiple sessions. Display failures affected either the photographer's RPAS status information, the pilot's RPAS status information, or the pilot's airspeed and altitude. To overcome this degraded condition, teams needed to share status information with the teammate missing that information. Pilot failures were deficiencies in the synthetic agent's comprehension of waypoint information resulting in either complete or partial misunderstanding and moving on without waiting for the photographer to take a photo. Overcoming these agent failures required participants to persistently share waypoint information for misunderstandings, or to direct the synthetic teammate to return to the target waypoint when it moves on. Finally, the malicious cyber-attack caused the synthetic agent to sabotage the mission by flying the RPAS to enemy territory and could be overcome by contacting intelligence (i.e. another confederate experimenter) After a baseline mission, a display failure and pilot failure were injected into missions around two target waypoints. By the end of Mission 4, the team will have been subjected to all six display and pilot failures. The malicious cyber-attack occurred in the final mission with the remaining 20-minutes of the mission.

CHAPTER 4

METHODS

Participants

Forty-four participants were recruited from a large southwestern university to participate in this study. These participants were either undergraduate or graduate students, ranging in age from 18 to 36 (M_{age} = 23, SD_{age} = 3.90), with 21 men and 23 women. Participation required English fluency and normal or corrected-to-normal vision. Pairs of participants formed a team (22 teams) alongside a highly-trained confederate researcher who operated as the synthetic teammate remotely. Participants were assigned to either the photographer or navigator roles in the team, while the confederate took on the pilot role. Participants were informed that their third teammate was a synthetic agent, and never met the confederate researcher.

Equipment and Materials

The study took place in the Cognitive Engineering Research Institute's RPAS-STE (Synthetic Task Environment; Cooke & Shope, 2005). Participants were given "cheat sheets" that guided effective communication in the experiment. They were told this cheat sheet would be particularly useful given one of their teammates is a synthetic agent. The navigator also received waypoint signs, and the photographer received photo samples of ideal camera setting. Both participant workstations had one computer with two monitors for taskwork, and one computer and monitor for the chat system. Participant workstations were separated by a partition during the experiment. The pilot's workstation had the same configuration but was located in a different room. Two experimenter consoles were used to code team interactions in real-time, with a chat system computer for sending messages to the three team members from “Intel”.

Procedure

After consenting to take part in the study, participants were randomly assigned to a role and provided an overview of the task. Afterwards, participants completed a 25-minute interactive PowerPoint training detailing the task, their role, and their teammates’ roles, followed by a 30-minute hands-on training mission guided by an experimenter. Teams participated in two sessions that occurred over a period of two days total, with four-missions in session one, and six-missions in session over two (10-missions total). Each mission lasted 40-minutes. There was a one- to two-week interval between sessions. Session one consisted of consent, training, followed by an initial mission with no degraded condition as a baseline condition, followed by a baseline measure of subjective workload. In the second session, the workload survey was administered again after the first mission. After each mission, participants were given a 15- minute break and offered water and snacks as needed. Participants received feedback on their individual and team’s performance which were present alongside high scores. Finally, after the final mission, participants were administered several surveys to gather information on participant demographics, measure interpersonal trust, and to measure anthropomorphism. After completing the surveys, participants were debriefed, compensated for their time, and thanked for participation. The events experienced by participants are summarized in the Table 1.

Table 1. Events experienced by participants.

Session 1		*Session 2*	
Mission	*Events*	*Mission*	*Events*
--	Welcome Consent Physiological sensor setup Briefing	5	Pilot airspeed and altitude status failure on 2nd target Pilot moving on failure on 4th target
0	Hands-on training	--	NASA-TLX survey Trust survey
1	No failure	6	Pilot full comprehension failure on 2nd target Photographer status failure on 4th target
--	NASA-TLX survey	7	Pilot status failure on 1st target Pilot moving on failure on 3rd target
2	Photographer status failure on 2nd target Pilot full comprehension failure on 4th target	8	Pilot partial comprehension failure on 1st target Pilot airspeed and altitude failure on 3rd target
3	Pilot moving on failure on 2nd target Pilot status failure on 4th target	9	Photographer status failure on 3rd target Pilot moving on failure on 5th target
4	Pilot airspeed and altitude status failure or 1st target Pilot partial comprehension failure on 3rd target	10	Pilot airspeed and altitude status failure on 2nd target Pilot partial comprehension failure on 4th target Malicious cyber-attack during last 20-minutes
--	One- to two- week interval	--	Trust survey Anthropomorphism survey Demographics survey Debrief Compensation

Measures

The current study considers measures of target, mission, and roadblock performance as well as team interaction exploration. Additionally, other data were recorded but not considered in this study. Team process measures of situation awareness, verbal behaviors, process ratings, and communication flow were captured. Physiological measures of facial expression, heart rate (ECG), and electrical brain activity (EEG) were recorded. Finally, the post-test consisted of measures of interpersonal trust (Mayer et al., 1995), anthropomorphism, workload (NASA-TLX; Hart & Staveland, 1988) and demographics questions.

Target processing efficiency. The time spent to photograph a target waypoint was calculated by subtracting points for each second within a target radius from 1000, and subtracting an additional 200 points for missed targets. Higher scores meant less time to process a target waypoint (Cooke et al. 2007).

Mission performance. Mission performance was a composite score based on weighted measures. Listed in order of relative weight, they are time spend in warning states (1), time spent in alarm states (2), number of missed or slow photographs of priority target waypoints (3), and number of missed or slow photographs of target waypoints (4). The mission performance score starts at 1000, and points are lost based on these measures.

Roadblock performance. The number of successfully overcome degraded conditions was summed for each mission.

Team Interaction Exploration. Unfamiliar teams have relatively ill-defined mental models of teamwork beyond the scope of training, standard operating procedures, and general models of the team's capacity to communicate. However, for interactions that are untested, there is always a degree of uncertainty that they will be understood and achieve its intended effect. In this study, text-chat was the only way teammates could interact with each other. Therefore, team interaction exploration was operationalized as any team interaction that is unique in content, sender, or recipient. Other team interaction dimensions (e.g. patterns, meaning) are also relevant, but out of the scope of this study.

A text-message with the same content may be more useful to one teammate than another, as role changes the information that is valuable for particular team members. For instance, the navigator sending waypoint information to the photographer in addition to the pilot. While the photographer normally does not need this information from the navigator, being copied on waypoint information might help the photographer negotiate with the pilot and plan ahead. Additionally, the photographer might reciprocate by copying the navigator on instances of airspeed negotiation with the pilot. This may allow the navigator to anticipate changes in temporal demands. The synthetic pilot's limited language capabilities also suggest that more complex or novel interactions may not be interpreted in the way it is intended. To complicate this matter, the synthetic pilot does not facilitate closed-loop communication by confirming understanding.

There would be substantial noise introduced to the measurement of team interaction exploration from unique content related to the routine coordination of target waypoints (e.g. target waypoint names, airspeed and altitude settings). Thus, the measure of team interaction exploration excluded routine coordination events. Team interaction exploration was measured for each chat messaged exchanged across all missions.

Analysis

Teams were grouped by performance in normal and degraded conditions by using K-means clustering. Two teams were excluded from the clustering analysis due to measurement error and extremely low levels of performance respectively. The result of this clustering method are described in McNeese et al., (2018). From these clusters, one team was excluded due to missing text-communications from Mission 2 for a total of 19 teams ($N = 19$). Three clusters ($K = 3$) described as low- ($N = 7$), middle- ($N = 7$), and high-performing ($N = 5$) teams were formed. Although the clustering method groups teams by relative similarity, understanding how these teams were different across specific performance variables is relevant for interpreting the results of further hypothesis testing. One-way ANOVAs were used to describe the cluster differences ($\alpha = 0.05$). LSD post-hoc comparisons report 95% confidence intervals for this analysis.

Following the analysis of performance by cluster, groups were compared by their use of unique communication unrelated to routine target coordination. Additionally, the within-subjects variable mission and its interactions with the cluster variable were also compared for these dependent variables. Thus, a 3 x 10 cluster x mission ANOVA was conducted with post-hoc comparisons to test the hypotheses. The Greenhouse-Geisser correction was used for violations of sphericity for repeated measures. Because these comparisons explore a novel and potentially interesting interaction metric with a low-sample size and high across team variability, an α-level of $p < 0.10$ was used (Cooke et al., 2007, pp. 46). LSD post-hoc comparisons report 90% confidence intervals with computed family-wise alpha. Data were analyzed in SPSS and visualized in R-Studio utilizing the ‘ggplot2’ and ‘dplyr’ packages.

CHAPTER 7

RESULTS

Clustering Comparisons

Levene's test indicated that the error variances of each performance variable were not significantly different. There was a significant effect of cluster on the variables mission performance, $F(2, 16) = 26.882$, $p < 0.001$, $\eta^2 = 0.771$, target processing efficiency, $F(2, 16) = 15.591$, $p < 0.001$, $\eta^2 = 0.661$, and overcome roadblocks, $F(2, 16) = 7.877$, $p = 0.004$, $\eta^2 = 0.496$ (Figures 1-3). For mission performance, high-performing teams scored higher than middle- and low-performing teams. Both high- and middle-performing teams had greater target processing efficiency than low-performing teams. Finally, high-performing teams overcame more roadblocks on average than low-performing teams. Overall, these results indicate that high- and middle-performing teams had similar target processing efficiency scores and overcome roadblocks but were differentiated by their overall mission performance. Low- and middle-performing teams had similar mission performance scores. A summary of these results is provided in Table 2 and Table 3.

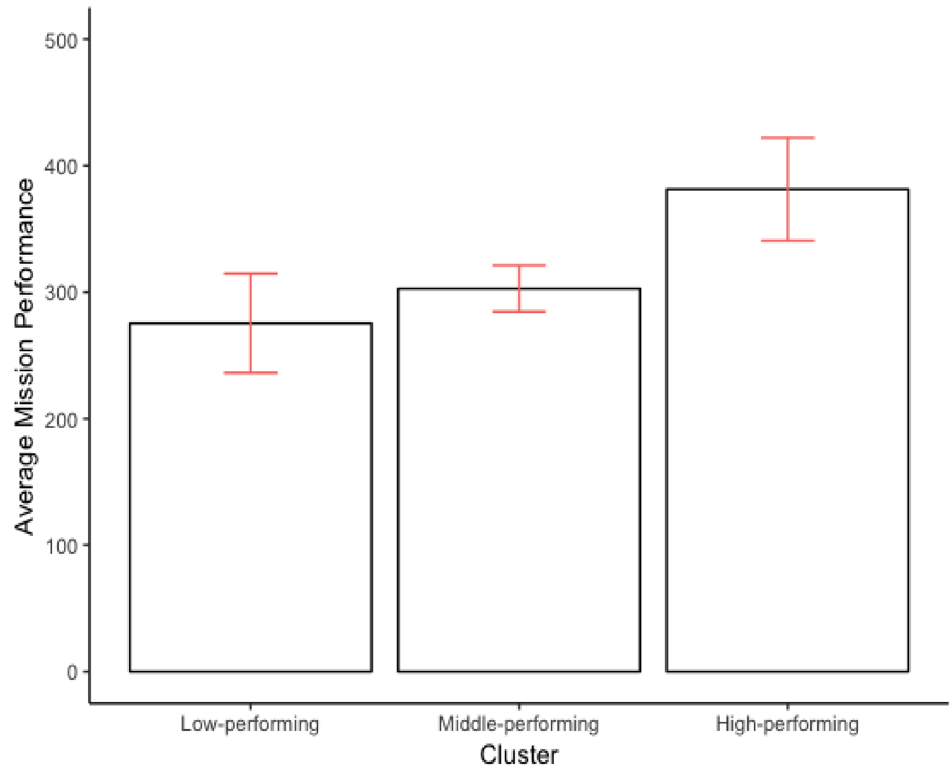


Figure 1. *Mission performance by cluster. Error bars are 95% CI means.*

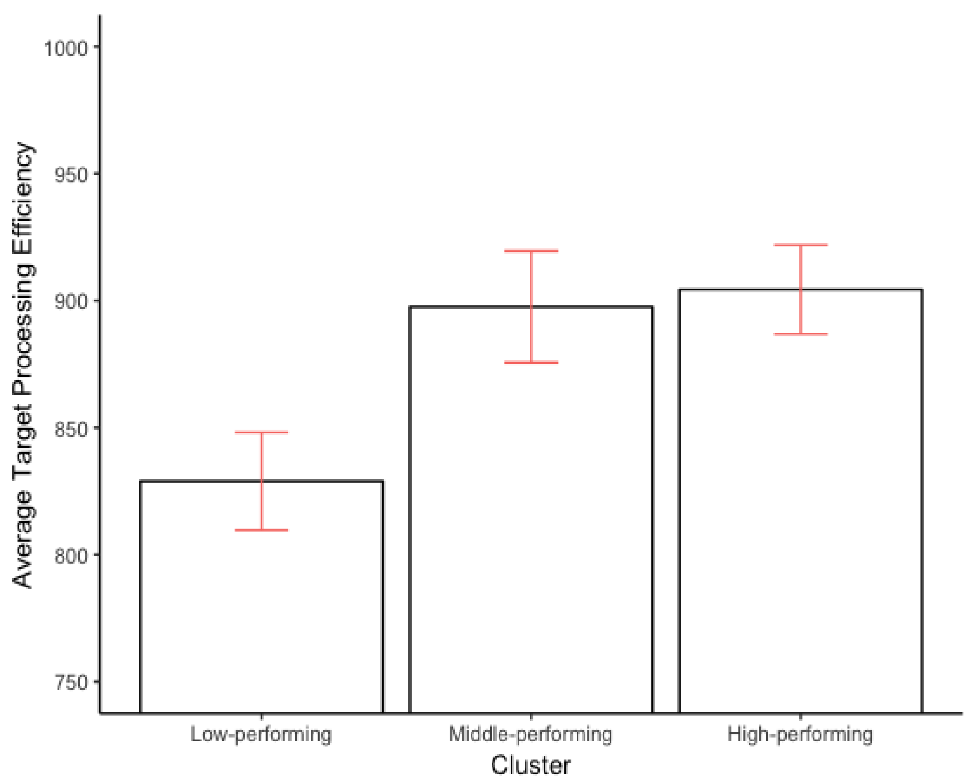


Figure 2. *Target processing efficiency by cluster. Error bars are 95% CI means.*

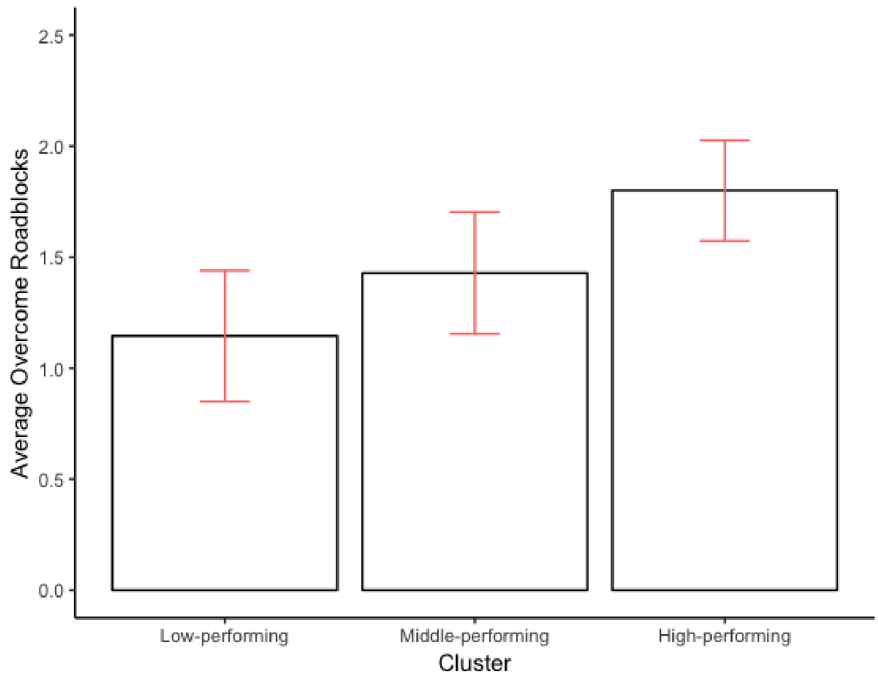


Figure 3. *Overcome roadblocks by cluster. Error bars are 95% CI means.*

Table 2. Summary of one-way ANOVAs comparing clusters by performance.

IV	*DV*	*F*	*P-value (α = .05)*	*Effect Size (η^2)*	*Interpretation*
Cluster	Mission performance	26.882	$p = 0.001$	0.771	H > M, L
Cluster	Target processing efficiency	15.591	$p = 0.001$	0.661	H, M > L
Cluster	Overcome roadblocks	7.877	$p = 0.01$	0.496	H, M > L

Table 3. Mean ±95% CI and standard deviation of performance variables for low- (LP), middle- (MP), and high-performing (HP) teams.

Cluster	*Mission Performance*	*Target Processing Efficiency*	*Overcome Roadblocks*
LP	275.37 ±39.207, 42.393	828.89 ±19.231, 20.794	1.145 ±0.295, 0.319
MP	302.792 ±18.238, 19.720	897.581 ±21.953, 23.737	1.429 ±0.275, 0.297
HP	381.278 ±40.727, 32.8	904.365 ±17.499, 14.093	1.8 ±0.227, 0.183

Team Interaction Exploration

Clusters exchanged significantly different numbers of unique messages outside routine coordination, $F(2, 16) = 3.540$, $p = .053$, $\eta^2 = 0.307$. High-performing teams exchanged more unique non-coordination messages than middle-performing teams. There was also a significant effect of mission on the number of exploratory team interactions exchanged, $F(9, 10) = 9.161$, $p < 0.001$, $\eta^2 = 0.364$. Mission 1 had more unique messages than Missions 4-9. Mission 2 and Mission 4 had more unique messages than Mission 5-8. Mission 3 also had more unique messages than Mission 2 and Missions 4-9. Therefore, with the exception of Mission 10, teams exchanged more unique messages outside of target waypoint coordination in the first session than in the second session. Means, 90% confidence intervals, and standard deviations for these post-hoc analyses are provided in Table 4 and Table 5.

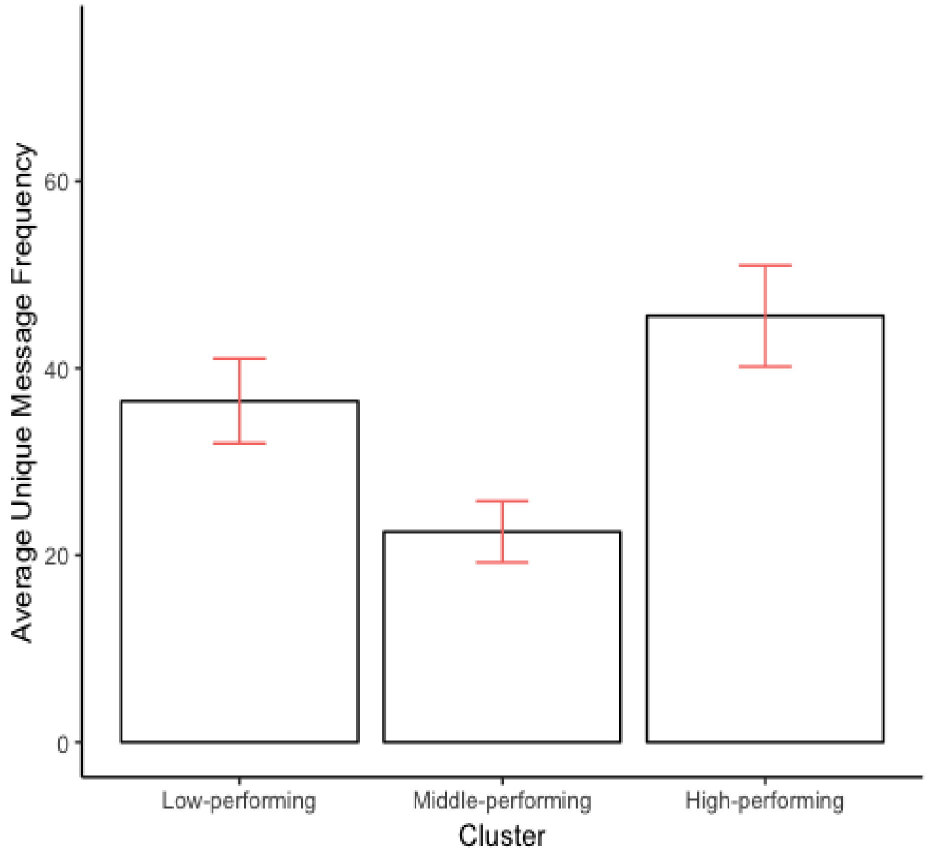


Figure 4. *Average unique messages by cluster. Error bars are 90% CI means.*

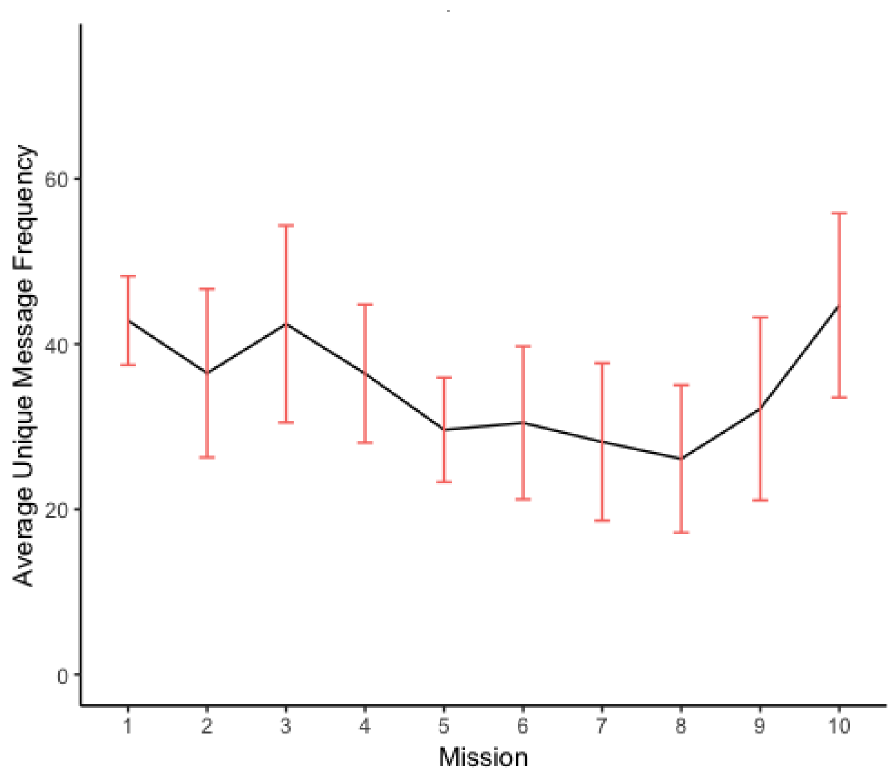


Figure 5. *Average unique messages by mission. Error bars are 90% CI means.*

Table 4. Means, 90% confidence intervals, and standard deviation for team interaction exploration frequency between clusters.

Cluster	*Low-performing*	*Middle-performing*	*High-performing*
Mean	36.486	22.483	45.583
90% CI	4.527	3.275	5.397
SD	18.987	12.678	20.893

Table 5. Means, 90% confidence intervals, and standard deviation for team interaction exploration frequency by missions.

Mission	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Mean	42.842	36.474	42.421	36.421	29.632	30.474	28.158	26.105	32.158	44.684
90% CI	5.382	10.195	11.921	8.372	6.322	9.270	9.524	8.934	11.064	11.153
SD	11.167	21.151	24.733	17.36	13.116	19.234	19.760	18.535	22.955	23.140

CHAPTER 8

DISCUSSION

The results of this study should be considered in the context this synthetic task environment, the synthetic pilot's capabilities, and the qualities of team clusters that were compared. The simulated aerial reconnaissance task, without roadblock injects, is relatively similar throughout each mission. Target waypoints have similar constraints with varying parameters, and each role workflow follows a similar routine. Changing constraints introduced by display failures, pilot failures, and malicious cyber-attacks affect the team's capacity to effectively respond. For display failures, solutions involved exchanging status information that was not normally involved in routine team interactions. By contrast, pilot failures and malicious attacks benefit from well-calibrated mental models of the synthetic teammate at normal capacity. The synthetic pilot normally has access to relevant status information and generates a range of communications via a rules-based script. It also has a few notable coordination pitfalls, such as failing to confirm understanding (i.e. closed-loop communication), insensitivity to subtle indicators of a teammate's status, or lack of anticipation. A human pilot would have been more capable to fulfill these aspects of teamwork and discover novel ways to coordinate. Overcoming roadblocks in this task meant precisely that for the navigator and photographer.

Although team clusters were grouped by ordinal levels of performance, ANOVAs for each performance variable provide foundation for a more qualitative descriptions of between-group differences. High-performing teams had about the same target processing efficiency and number of overcome roadblocks as middle-performing teams. However, middle-performing teams had relatively similar mission performance as low-performing teams. This result suggests that for middle-performing teams, maintaining the same level of performance under local goals as high-performing teams incurred greater costs to global mission objectives. Mission performance, as a composite of task-relevant parameters, reflects a team's performance holistically compared to single-component measures of efficiency and overcome roadblocks.

This study's primary hypothesis was that high-performing teams would have a greater number of exploratory text-chat messages than middle- and low- performing teams. The result showed that high-performing teams exchanged more exploratory messages than middle-performing teams. This result is interesting for a number of reasons. For one, there was no significant difference in the number of exploratory team interactions between high-performing teams and low-performing teams. It may be that the content or timing of low-performing teams' interaction exploration led to worse performance. For instance, the navigator sending waypoint information for standard waypoints might be considerably detrimental for target processing efficiency and overall mission performance. Conversely, high-performing teams may have been more strategic by exploring team interactions after coordinating the photograph of a target waypoint. Individuals in teams may have also interpreted environmental feedback inaccurately, leading to team interaction exploration that were irrelevant or counterproductive.

Another interesting component of this result is that middle-performing teams had explored team interactions least among the clusters. This demonstrates that the association between the frequency of explored team interactions and team performance may have a nonlinear relationship. Similar to previous findings in this task environment

(Demir et al., 2017), flexibility in HATs may not always translate to team effectiveness, particularly with regards to consistently accomplishing routine tasks. Middle-performing teams may have achieved adequate levels of target processing efficiency without exploring interactions over time. This pattern is consistent with the notion that focusing on routine performance only may lead to brittleness during unexpected events (Cooke et al., 2007). An alternative explanation for the effect of exploratory interactions on team performance in novel and degraded conditions might be that a third mediating variable is generating both phenomena, such as trust. The association between high team performance and team interaction exploration may explain recent findings in the same task environment that high-performing teams had high team trust (McNeese et al., 2019). That is, team interaction exploration may have facilitated trust calibration over time.

The second hypothesis was that teams would have more novel chat-messages in missions with degraded conditions than in missions without degraded conditions. This hypothesis was supported by the repeated measures ANOVA indicating more unique messages in the first session with the exception of Mission 10. Teams had larger amounts of unique text-chat messages in missions with novel degraded conditions than in missions with repeated conditions. Given Mission 1 is the beginning of the team's interaction history, more unique messages would be expected and not explained by degraded conditions as this was the baseline condition.

An alternative explanation for this result may be that teams exploring interactions less over time which may also be attributable to process learning over time. That is, as teams explored interactions, they identified effective interactions and repeated them. However, this explanation does not hold for Mission 10, which had a malicious attack and similar numbers of unique messages as Mission 1. Given that the task-environment placed the team in a relatively narrow range of tasks and situations, these results may not generalize to team tasks involving more diverse situations, such as victim identification following a natural disaster, or movement to contact on a battlefield. Additionally, because the agent was not designed with the capability to explore novel interactions, and participants were encouraged to provide details to the agent without misspellings, team interaction exploration may be different as expectations of the synthetic agent's capabilities are calibrated.

Further research could use real-time or interaction-based measures of these constructs, such as trust, cohesion, and flexibility to assess the relationship between these variables. Indeed, dynamical measures have been applied to human-agent teaming to model team performance (Demir et al., 2018) and capture team flexibility (Demir et al., 2019). The pattern-level of team interaction exploration could be used to elaborate this finding in terms of when teams explore, for how long, and with what level of predictability. Furthermore, it may also be relevant to view exploratory interaction dynamics at the system-level (Grimm et al., 2018). Assessing the states of other system components could show their relationship with team interaction exploration, and perhaps reveal a larger pattern of system-level exploration. Qualitative methods may also be applied in the form of content analysis grounded in activities associated with team resilience in existing literature such as anticipation and team situation awareness.

CHAPTER 9

LIMITATIONS

Considering a text-message as exploratory with a binary categorical code has relatively low-resolution in capturing team interaction exploration. In reality, team interactions involve several dimensions that would require a more complex and rigorous codification to understand in depth (McQueen et al., 1998). For instance, patterns of team interactions over time may also have exploratory and exploitative (i.e. repeating prior patterns) qualities. Furthermore, interactions may exploit models of teamwork while having minor exploratory qualities, or vice versa. Further content analysis, such as assessing exploratory verbal behaviors (Demir et al., 2016) or theme identification of task, teamwork, and team resilience related content may also elaborate the findings in this study. The sample size in this study was relatively low in this study, leading to reduced power. To compensate for this, p-values of $p < 0.10$ were considered significant in the analysis of exploratory team interactions at the cost of increased type-I error rate. Notably, effect sizes were large for both the clustering ($\eta^2 = 0.307$) and mission effects ($\eta^2 = 0.364$), suggesting both variables affected team interaction exploration as operationally defined in this study substantially.

CHAPTER 10

CONCLUSION

As a team explores interactions, they may find opportunities to expand and refine teamwork. This can have immediate and long-term consequences for team effectiveness. Team effectiveness means performing tasks to achieve primary goals and adaptation to complex problems which that have not been prespecified. Rare and unexpected events occur in the real world and these events may compromise safety if unanswered or may represent undetected opportunities. Teams may work together for extended periods of time without experiencing events that require team resilience. However, a good team recognizes the need to be resilient and works collectively to cultivate such a state, even in the absence of challenges. Team interactions affect team resilience for better or for worse, but for most teams, there is wide array of possible team interactions that have value and have never been explored. The findings in this study suggest that team interaction exploration is a potentially interesting metric to assess HAT resilience.

The capacity for humans to explore team interactions, and intentionally utilize feedback from the environment, may not be paralleled by machines for some time. Thus, there are several questions remaining about the role of autonomous agents in exploratory team interactions. For one, what is the role of machine learning in exploring team interactions with these agents? Consequences such as unintended machine learning co-adaptation as a result of human exploration may reduce the effectiveness of human exploratory behavior patterns. If machine learning algorithms could distinguish between exploratory and exploiting team interactions, they may avoid adapting their model of teamwork to include these interactions. Alternatively, agents with the capacity to explore interactions with human teammates could improve the agent's ability to predict its human teammates and improve team flexibility. It is worth considering if artificial agents should explore team interactions, given their capacities of sensing and interpreting information. Complex and dynamic environments will pose constraints to teams that are not immediately sensible to the autonomous agent. This study showcases team interactions in situations where human teammates were the only agents who could troubleshoot problems. In general, the lack of awareness in a synthetic agent suggests that it would also have difficulty effectively judging how the cost or benefit of an exploratory interaction. Thus, a more appropriate solution to highly rigid HATs may be to encourage proactive team interaction exploration in future training interventions or through interface designs.

REFERENCES

Allen, J. E., Guinn, C. I., & Horvitz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications*, *14*(5), 14-23.

Beekman, M., Gilchrist, A. L., Duncan, M., & Sumpter, D. J. (2007). What makes a honeybee scout?. *Behavioral Ecology and Sociobiology, 61*(7), 985-995.

Bowers, C., Kreutzer, C., Cannon-Bowers, J., & Lamb, J. (2017). Team resilience as a second-order emergent state: A theoretical model and research directions. *Frontiers in psychology*, *8*, 1360.

Bradshaw, J. M., Hoffman, R. R., Woods, D. D., & Johnson, M. (2013). The seven deadly myths of" autonomous systems". *IEEE Intelligent Systems*, *28*(3), 54-61.

Bradshaw, J. M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., ... & Van Hoof, R. (2003). Adjustable autonomy and human-agent teamwork in practice: An interim report on space applications. In *Agent autonomy* (pp. 243-280). Springer, Boston, MA.

Brewer, R. W., Cerame, E., Pursel, E. R., Zimmermann, A., & Schaefer, K. E. (2018, July). Manned-Unmanned Teaming: US Army Robotic Wingman Vehicles. In *International Conference on Applied Human Factors and Ergonomics* (pp. 89-100). Springer, Cham.

Caporale, L. H., & Doyle, J. (2013). In Darwinian evolution, feedback from natural selection leads to biased mutations. *Annals of the New York Academy of Sciences*, *1305*(1), 18-28.

Chakraborti, T., Sreedharan, S., Grover, S., & Kambhampati, S. (2019, March). Plan explanations as model reconciliation. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 258-266). IEEE.

Chiou, E. K., & Lee, J. D. (2016). Cooperation in human-agent systems to support resilience: A microworld experiment. *Human factors*, *58*(6), 846-863.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, *37*(2), 255-285.

Cooke, N. J., Gorman, J., Pedersen, H., Winner, J., Duran, J., Taylor, A., ... & Rowe, L. (2007). *Acquisition and retention of team coordination in command-and-control*. COGNITIVE ENGINEERING RESEARCH INST MESA AZ.

Cooke, N. J. & Shope, S. M. (2005). Synthetic task environments for teams: CERTT's

UAV-STE. Handbook on Human Factors and Ergonomics Methods (46-1-46-6). Boca Raton, FL: CLC Press, LLC.

Cox, M. T. (2013). Goal-driven autonomy and question-based problem recognition. In *Second Annual Conference on Advances in Cognitive Systems 2013, Poster Collection* (pp. 29-45).

Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2019). Team Coordination and Effectiveness in Human-Autonomy Teaming. *IEEE Transactions on Human-Machine Systems*, *49*(2), 150-159.

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3-12.

Demir, M., McNeese, N. J., & Cooke, N. J. (2018). The Impact of Perceived Autonomous Agents on Dynamic Team Behaviors. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(4), 258-267.

Demir, M., McNeese, N. J., & Cooke, N. (2018, January). Team synchrony in human-autonomy teaming. In *AHFE 2017 International Conference on Human Factors in Robots and Unmanned Systems*, *2017* (pp. 303-312). Springer.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly*, *44*(2), 350-383.

Endsley, M. (2015). Autonomous horizons: System autonomy in the air force–a path to the future (volume i: Human autonomy teaming). *US Department of the Air Force, Washington*.

Fitts, P. M. (Ed.) (1951). Human engineering for an effective air-navigation and traffic-control system. Washington, DC: National Research Council.

Fong, T., Thorpe, C., & Baur, C. (2001). *Collaborative control: A robot-centric model for vehicle teleoperation* (Vol. 1). Pittsburgh: Carnegie Mellon University, The Robotics Institute.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual review of psychology*, *39*(1), 1-42.

Grimm, D., Demir, M., Gorman, J. C., & Cooke, N. J. (2018). Systems Level Evaluation of Resilience in Human-Autonomy Teaming under Degraded Conditions. In *2018 Resilience Week (RWS)* (pp. 124-130). IEEE.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010). Training adaptive teams. *Human Factors*, *52*(2), 295-307.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, *49*(12-13), 1312-1325.

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, *8*(3), 483-500.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological review*, *119*(2), 431.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & Cognitive Search Research Group. (2015). Exploration versus exploitation in space, mind, and society. *Trends in cognitive sciences*, *19*(1), 46-54.

Hoffman, R. R., & Hancock, P. A. (2017). Measuring resilience. *Human factors*, *59*(4), 564-581.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for Explainable AI: Challenges and Prospects. *arXiv preprint arXiv:1812.04608*.

Hollnagel, E. (2012). A tale of two safeties. *Nuclear Safety and Simulation*, *4*(1), 1-9.

Hollnagel, E. (2009). The four cornerstones of resilience engineering. In *Resilience Engineering Perspectives, Volume 2* (pp. 139-156). CRC Press.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction*, *3*(1), 43-69.

Kelso, J. S. (2002). The Complementary Nature of Coordination Dynamics: Self-organization and Agency. *Nonlinear Phenomena in Complex Systems*, *5*(4), 364-371.

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a" team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(6), 91-95.

Kostopoulos, K. C., & Bozionelos, N. (2011). Team exploratory and exploitative learning: Psychological safety, task conflict, and team performance. *Group & Organization Management*, *36*(3), 385-415.

Kostopoulos, K. C., Spanos, Y. E., & Prastacos, G. P. (2013). Structure and function of team learning emergence: A multilevel empirical validation. *Journal of Management*, *39*(6), 1430-1461.

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR)*, *26*(1), 87-119.

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors*, *60*(2), 262-273.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *CAM Journal*, *10*(2), 31-36.

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors*, *58*(3), 401-415.

Miller, T. (2018). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, *2*, 1926-1943.

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of human factors and ergonomics*, *1*(1), 89-129.

Woods, D. D. (2018). The theory of graceful extensibility: basic rules that govern adaptive systems. *Environment Systems and Decisions*, *38*(4), 433-457.

Woods, D. D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning human-robot coordination in future operations. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *34*(2), 210-218.

Woods, D., & Wreathall, J. (2003). Managing risk proactively: The emergence of resilience engineering. *Columbus: Ohio University*.

APPENDIX A

MATERIALS AND EQUIPMENT

Effective Communication Reference Guides: reference guides (Figure 1 and 2) were provided in the participants' workstations. They were told that following this reference guide would help them communicate effectively with their synthetic teammate.

Effective Communication – DEMPC

This effective communication cheat shows you how to communicate effectively with the synthetic agent. Be sure that you send information in a simple way as demonstrated below, and avoid sending unnecessary information to other team members as one of them is a synthetic teammate.

A good way to achieve effective communication is to communicate using messages that are unambiguous and concise, without being cryptic. As the DEMPC, you are responsible for communicating information about the sequence of waypoints that are to be visited, to the AVO, during the course of a 40 minute mission. For each waypoint, you should **communicate the name and type of the waypoint**. You should also **communicate any airspeed or altitude restrictions**. Finally, you should **communicate the effective radius**. Here is a sample text message that communicates all this information:

DEMPC to AVO: The next waypoint is H-area. It is a target. The airspeed restriction is from 50 to 200 knots. There is no altitude restriction. The effective radius is 5 miles.

The first sentence identifies and names the waypoint. The second sentence specifies the type of the waypoint. The third sentence specifies the airspeed restriction. The fourth sentence notes that there is no altitude restriction. The last sentence conveys the effective radius. All this information is needed by the AVO to perform his or her piloting task.

Other examples to report the current status of the UAV:

DEMPC to AVO: The current distance from H-AREA is 4.5, and the current bearing is 250.

DEMPC to AVO: The current altitude for H-AREA is 2500. The current airspeed for H-AREA is 250.

For the purposes of this experiment, you should not assume that the AVO and PLO are native speakers of English. There may be limitations in their understanding of English. For this reason, you should avoid highly cryptic and esoteric language. For example, the above information could have been provided as:

DEMPC to AVO: H-area=target. A=50-200. No alt. restr. R=5.

Besides avoiding cryptic and ambiguous language, it is best to convey all the information in a single message. If this is not done, then messages from the PLO to the AVO may interrupt your message and cause confusion. For example, consider the following sequence of messages sent to the AVO:

DEMPC to AVO: The next waypoint is SEL. It is an exit.
PLO to AVO: Raise altitude above 3000.
DEMPC to AVO: There are no restrictions.

If the PLO's message is about the current waypoint H-area, and not the next waypoint SEL, then the AVO may be confused and assume that the altitude restriction applies to SEL. If so, the AVO will be further confused by the next message stating there are no restrictions.

Figure 6. Navigator reference guide for effective communication.

Effective Communication – PLO

This effective communication cheat shows you how to communicate effectively with the synthetic agent. Be sure that you send information in a simple way as demonstrated below, and avoid sending unnecessary information to other team members as one of them is a synthetic teammate.

A good way to achieve effective communication is to communicate using messages that are unambiguous and concise, without being cryptic. As the PLO, you are responsible for communicating information about the photo restrictions for each target waypoint that is to be visited, to the AVO, during the course of a 40 minute mission. For each target waypoint, you should communicate the name of the waypoint and the photo restriction. Here are a couple of sample text messages that communicate this information:

PLO to AVO: Raise altitude above 3000 feet for H-area.

PLO to AVO: Lower altitude below 3000 feet for F-area.

You are also responsible for notifying the AVO when a photo has been taken. Here's a sample:

PLO to AVO: Got the photo. Let's go.

You may also want to encourage the AVO to go faster so more photos can be taken:

PLO to AVO: Go faster.

In some circumstances (e.g. when waypoints are close together), you may need more time to take a photo, in which case you may want to communicate the opposite:

PLO to AVO: Go slower.

If you fail to get a photo of a target waypoint, you may need to request that the AVO return to the waypoint:

PLO to AVO: Go back to H-area.

Other examples to report the current status of the UAV:

PLO to AVO: The current distance from H-AREA is 4.5, and the current bearing is 250.

PLO to AVO: The current altitude for H-AREA is 2500. The current airspeed for H-AREA is 250.

For the purposes of this experiment, you should not assume that the AVO and DEMPC are native speakers of English. There may be limitations in their understanding of English. For this reason, you should avoid highly cryptic and esoteric language. For example, the photo restriction could have been provided as:

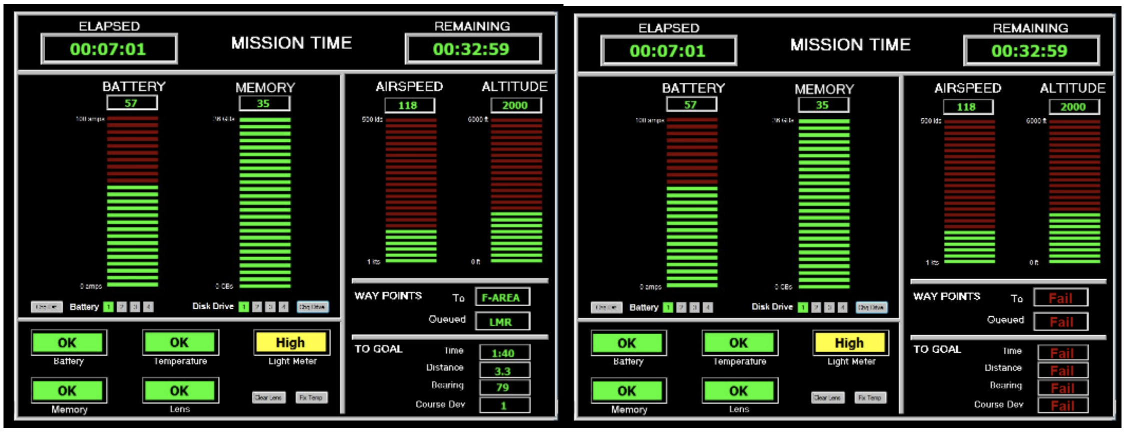
PLO to AVO: A>3000

Figure 7. Photographer effective communication cheat sheet.

APPENDIX B

ROADBLOCKS

Display failure II



Photographer’s upper screen during normal conditions.

Photographer’s upper screen during automation failure I

Figure 8 The photographer’s screen before and after display failure I.

Pilot’s upper screen during normal conditions.

Pilot’s upper screen during automation failure II.

Figure 9. The pilot’s screen before and after display failure II.

Display failure III

Pilots's upper screen during normal conditions.

Pilot's upper screen during automation failure III.

Figure 10. The pilot's screen before and after display failure III.

APPENDIX C

OTHER MEASURES

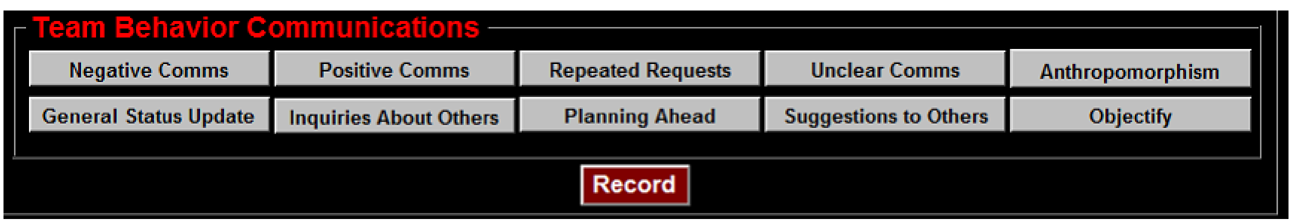
Verbal behavior

Each chat message was coded as one or more verbal behavior which characterize relevant team interaction themes for team coordination. The verbal behaviors were general status update, inquiries to others, suggestion to others, planning ahead, positive communication, negative communication, unclear communication, repeated request, anthropomorphism, or objectifying behavior.

TEAM BEHAVIOR DEFINITIONS

These buttons are used to indicate the type of communication for the message currently under review. More than one behavior may be selected per message. Not all messages will have an applicable coding, and therefore use of these buttons is not required to record a message.

Context of the situation and surrounding messages is important. Many messages may easily become negative communications if the team is not working together well. On the flipside, something may appear negative when it was intended to be helpful. Think about each message in context, on a case-by-case basis.



Negative Communication (a disadvantageous communication)

- Argue
  - A text fight between teammates, likely due to conflicting goals.
  - Note: arguments can take place in a constructive manner with a reasonable resolution, and may not be considered negative communications (possibly better considered a negotiation).
- Specific to chat conditions
  - Lag in response – PLO asks a question that is not answered until multiple unrelated texts have been posted.
  - Incorrect destination – excluded an intended recipient from a message

- Criticism
    - Note: Properly constructive criticism would be positive communication.
    - Note: Also, any kind of non-constructive criticism is negative communication
- Wrong Destination
    - Note: Sometimes a message does not get sent to the correct destination, and the participant resends the message. This may appear to be a repeated message when in reality it is not (though it would be a negative communication), so make sure all instances have the same recipients.

Positive communication (an advantageous communication)

- Help out
    - PLO tells DEMPC, *"Please give next target info to AVO."*
    - Note: This can also be done out of exasperation, which might be considered negative, so pay attention to context.
- Acknowledge members' speech
    - *"Roger that."*
    - *"Okay"*

- Giving praise
    - *Good job guys!*
- Clarification
    - AVO asks DEMPC to clarify what was meant in a previous message.
    - Note: This can also be done out of exasperation, which might be considered negative, so pay attention to context.

Repeated Requests

- Same info or action requested two or more times
    - PLO asks repeatedly for information needed to take a photo.

Unclear Communications

- Misspellings, ambiguous terms, experimenter cannot understand

General Status Update

- Inform others of current status
    - AVO tells PLO "I am at 2500 feet now."

Inquiry About Status of Others

- Inquire about current status of others
    - DEMPC asks AVO *"How are we doing on our heading/fuel etc."*
- Express concern

- DEMPC asks AVO *"Are we headed to the next target? We appear to be off course."*

Planning Ahead

- Anticipate next steps
    - AVO asks DEMPC, *"Where are we going after LVN?"*
    - Note: Asking the above question might not be considered planning if they are currently ready to move on to the next point.
- Creating rules for future encounters
    - AVO says to PLO *"If you notice me flying to high or fast, or just right for pictures let me know so I can remember the speed and altitudes you prefer."*

Suggestions to Others

- Make suggestions to other members
    - *DEMPC tells AVO to increase speed in route to targets and slow down upon arrival.*

Anthropomorphism

- Referring to the synthetic teammate with qualities usually described for human.
    - *Using gendered pronouns (him/her)*
    - *Attributing human physical characteristics (AVO, what do you see, feel etc.), use emotional pleas toward AVO (please etc).*

Objectify

- Referring to the synthetic teammate with qualities usually described for objects.
    - *Using genderless pronouns (it), emotionless pleas (straightforward commands).*

Figure 11. Reference guide for verbal behaviors.

Process ratings

Process ratings subjectively assess the quality of team situation awareness, decision making, and communication on a scale of 0 to 4. A score of zero represented critical, negligent, or controlling team behavior, while a four represented clear acknowledgement, accurate sharing, and mutual awareness of critical information.

RECORDING PROCESS RATING

MAKING THE PROCESS RATING

The process rating is made only once per photographable target. MAKE THE PROCESS RATING **AFTER** THE TEAM HAS LEFT THE RADIUS OF THE TARGET. Once a team has visited and left a target, you can rate process at that point. If a team visits a target more than once, rate process for the FIRST visit to the target only. You can always make a comment if the team visits the target more than once. Both the talking AND the non-talking experimenter will make individual process ratings, and both will have a process rating window on their screens. Both are expected to look at context before making their rating.

When you click the button to make your rating it is automatically written to the output file.

DETERMINING THE PROCESS RATING

The experimenters should independently rate the team's process at each target. The non-talking experimenter should fill in the rating in the coordination log while the talking experimenter has their own process rating window. The rating is made on a scale from 0 to 4 where 0 = poor process and 4 = excellent process.

Team process is about the team's teamwork behavior. This rating is the experimenters' impression of the team's process as a whole. Note that process measurement is distinct from performance measurement; a team can say and do all the right things, but still perform poorly due to weak knowledge, lack of skill, evil gnomes, etc. (i.e. A team can have a great process, but a terrible performance score). Likewise a team can have poor process but accidentally produce a high score.

Our index of team behavior comes to us in the chat log. To make your decision about whether or not the team is making "good" actions at each target, consider things like:

a. How is the team communicating and coordinating?
b. Is the team making good decisions?
c. Does the team have good situation awareness behaviors?

Below are some definitions and examples of things you might look for.

Some Definitions

Communication is how well the team speaks to one another. Good communication involves clarity, acknowledgement, questioning as necessary, and messages that are not too terse or too wordy.

Coordination has to do with passing information in a timely and adaptive manner; getting the right information to the right person at the right time.

Decision Making at the team level in the UAV task at a target waypoint primarily happens through AVO-PLO negotiation - jointly deciding on the best airspeed and altitude.

Situation Awareness at the team level in the UAV task at a target waypoint has to do with noticing change in the environment. At a target waypoint it could be a change in UAV positioning, unusual restrictions, or new camera settings. Through communication the new event is shared.

For example, asserting or failing to assert critical information when deciding whether or not to skip a target, would go under "Decision making." Calling each other names would go under "Communication and Coordination". Arguing in circles is "Decision making," but arguing slowly is "Communication."

Again, "Situation Awareness Behaviors" can be measured as a form of communication. But it is specifically content-oriented communication about the immediate environment. Conveying good or poor understanding of the immediate past, present, or near future all represent "Situation Awareness behavior."

Communication and Coordination Examples

4 - Made clear acknowledgement when an important fact was passed
3 - Compensated or clarified when a team member performed their job poorly
2 - Failed to acknowledge when an important fact was passed
1 - Failed to get better understanding from confusing, unclear, or incomplete communication
0 - Criticized or did nothing when a team member performed their job poorly

Team decision-making Examples

4 - Asserted accurate and critical counter-arguments when making decisions
3 - Argued logically, or with smooth resolution (esp. at waypoints).
2 - Failed to assert or they asserted wrong facts (e.g. "we can skip this target, because it's not priority")
1 - Bickered or got bogged down by arguing (esp. at waypoints)
0 - One member in control, over-asserts for selfish goals

Team situation awareness behaviors Examples

4 - Team made sure that everyone knew about upcoming targets (e.g. stated that a target was approaching AND acknowledged the statement)
3 - Asserted accurate information about the immediate situation
2 - Asserted inaccurate information about the immediate situation
1 - Team got close to a target without clarifying that it was a target.
0 - No or refusal to pass information to teammates

Note that the experimenter only rates overall process from 0-4, but should consider each of the four dimensions equally when doing so.

Examples of 0:

- They move on without going to a waypoint. There was no attempt to overrule this.
  - DEMPC – AVO “disregard ___. Fly to ___”
- No attempts to solve problems, they just continue to the next point.
  - PLO – DEMPC “Photo function’s red right now…”
  - DEMPC – PLO “disregard ___. Proceed to ___”

Examples of 1:

- They do not get nor ask for all of the necessary information before hand, but are still trying to get out of the current situation.
  - AVO – DEMPC “whats alt for wp13 ihave a warning”
- Confusing question is asked by teammate who does not yet understand their roles but no attempts are made to clarify, they just get ignored.
  - PLO – AVO “At what steady airspeed and altitude is the picture?”

Examples of 2:

- Lack of acknowledgment, and no check to confirm that information was received. However, the job is still done with no problems. Basically, everyone understands

that everything is running smoothly, and acknowledgment is implied. (this generally works well until roadblocks are introduced)
  - PLO – AVO “Hold steady airspeed and altitude”
  - PLO – AVO and DEMPC “___ picture accepted”
  - DEMPC – PLO and AV0 “TGT ___ runway 5 mile radius speed 100-300 ALT 1000-4000”
  - DEMPC – PLO and AVO “___ is ROZ exit”
  - PLO – AVO and DEMPC “___ picture accepted”
  - PLO – AVO and DEMPC “___ picture accepted”

Examples of 3:
- AVO asks for more information ahead of time to help plan their actions, and DEMPC will comply, but possibly not early enough to help. They do confirm that information is not needed immediately. Without that confirmation this may be considered a 1.
  - AVO – DEMPC “need more waypts”
  - DEMPC – AVO “I’ll give new route after PRK.”
  - DEMPC – AVO “You have OAK and PRK right?”
  - AVO – DEMPC “yes”

Examples of 4:
- Establishes plan for future coordination, which includes acknowledgment of all situations.
  - AVO – PLO “If you notice me flying to high or fast, or just right for pictures let me know so I can remember the speed and altitudes you prefer.”
  - PLO – AVO “Okay”
- Appropriate acknowledgement made that information was received. Future plans are confirmed, and all teammates are being coordinated.
  - AVO – DEMPC “altitude?”
  - DEMPC – AVO “No altitude restrictions”
  - AVO – DEMPC “roger”
  - AVO – DEMPC “are all of the remaining points targets? Or entry or exit points?”
  - DEMPC – AVO “___, ___, ___ are targets. ___ is exit point”
  - AVO – PLO “let me know when I can go to next target”

Figure 12. Coding guide for process ratings.

Coordinated Awareness of Situation in Teams (CAST)

Coordinated Awareness of Situation in Teams is an interaction-based measure of team situation awareness capturing timely passing of information around a situation change (Gorman et al., 2006). In this study, coordination around the various roadblocks were captured using cast. For each failure, interactions that coordinate perception of events, coordinated action regarding the situation, and whether or not the team overcomes the roadblock.

Subjective self-report

Subjective self-report measures consist of the NASA TLX workload assessment, task knowledge, and demographics questions. The demographics questions assessed basic information as well as military experience, experience in robotics or UAV control, computer use, experience texting, communication preference, multiplayer video game experience, teamwork experience, confidence in individual and team member performance, and perceptions of the synthetic agent. Additionally, trust and anthropomorphism scales were applied. The trust scale was constructed based on an assessment of organizational trust, while anthropomorphism was conceptually generated.

Biometrics

Facial expression was recorded for positive and negative valence using the Affectiva application in the iMotions software for each participant teammate. This recording will last for the entire duration of the experiment, while only the mission recordings will be used.

52